

Europäisches Patentamt European Patent Office Office européen des brevets



(11)

EP 1 033 401 A2

(12)

EUROPEAN PATENT APPLICATION

(43) Date of publication: 06.09.2000 Bulletin 2000/36

- (51) Int CL7: **C12N 15/12**, C07K 14/47, C07K 16/18, G06F 17/30
- (21) Application number: 00200610.4
- (22) Date of filing: 21.02.2000
- (84) Designated Contracting States:

 AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU

 MC NL PT SE

 Designated Extension States:

 AL LT LV MK RO SI
- (30) Priority: 26.02,1999 US 122487 P
- (71) Applicant: GENSET 75008 Paris (FR)
- (72) Inventors:
 - Dumas Milne Edwards, Jean-Baptiste 75006 Paris (FR)
 - Duclert, Aymeric 94100 Saint-Maur (FR)

- Giordano, Jean-Yves 75018 Paris (FR)
- (74) Representative: Brasnett, Adrian Hugh et al Mewburn Ellis York House 23 Kingsway London WC2B 6HP (GB)

Remarks:

THE COMPLETE DOCUMENT INCLUDING REFERENCE TABLES AND THE SEQUENCE LISTING IS AVAILABLE ON CD-ROM FROM THE EUROPEAN PATENT OFFICE, VIENNA SUB-OFFICE.

(54) Expressed sequence tags and encoded human proteins

(57) The sequences of 5' ESTs derived from mRNAs encoding secreted proteins are disclosed. The 5' ESTs may be to obtain cDNAs and genomic DNAs corresponding to the 5' ESTs. The 5' ESTs may also be used

in diagnostic, forensic, gene therapy, and chromosome mapping procedures. Upstream regulatory sequences may also be obtained using the 5' ESTs. The 5' ESTs may also be used to design expression vectors and secretion vectors.

```
monkey, macaque, grast, bacteria, fruit fly, sea
These were derived from expressed sequence tags (ESTs)
finterest. They can be used in diagnostics,
spins, identification of mutations, to assess
r nutritional purposes. The present sequence is a
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            essed sequence tail secreted protein; cDNA isolation;
                                                                                                                                              and coding sequences of novel
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            HERLYLYLOGG OVYVELLE GOTTLY STERREKEYGTPHGILLINGY 369
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        MAGRICLWWLDLKSGAQAIWTELFWPHEKVDGALCMEKSLGPNSCSA 349
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            MAGRELWWLDLKSGAQATWTELFWPHEKVDGALCMEKSLGPNSCSA 420
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              ------16HGHRN 169
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          ALGERY YOLL USING ELREDEN MGENEPREPERENVEDYEMPOPGFIGHEN 240
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         VWKSHKWDRELISERWKNFPSPVDAAFRQGHNSVFLIKGDKVWVYP 120
                                                                                                                                                                                                                                                                                                                                                                                                     71; Gaps
                                                                                                                                                                                                                                                                                                                                                                                                                                           SECWSLAIATPLPPTSAHGNVAEGETRPDPDVTERCSDGWSFDATT 60
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 VSLCWSLAIAIPLPPISAHGNVAEGETKPDPDVTERCSDGWSFDAIT 60
                                                tor treatment of diseases, diagnostics, raising
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  MCSPHLVLSALTSDNESTYAFSGTHYWRLDTSRDGWHSWPIAHQ
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               MRCSPHLVLSALTSDNH: JATYAFSGTHYWRLDISRDGWHSWPIAHQ
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         WEEKLYLVOGTOVYVELTKAGYTLVSGYPKRLEKEVGTPHGIILDSV
                                                                                                                                                                                                                                                                                                                                                                Score 2139.5; DB 22; Length 462;
Pred. No. 1.1e-192;
                                                                                                                                                                                                                                                                                                                                                                                                        0; Indels
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    CYSDVEKLNAAKALPOFCNVTSLEGCTH 391
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            CYSEVEKLNAAKALPQPUNUTSLIJCTH 462
                                                                                                                                                                                                                                                                                                                                                                                                   0; Mismatches
                                                                                                                                                provides the protein
                                                                                                          1018; 1275pp: English
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          ents, SEQ ID NO: 4:85
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  Protein; 83 AA
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       cosome mapping.
                                                                                                                                                                                                                                                                                                                                                                  97.98;
84.68;
                                                                       arch use
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  entity)
                                                                                                                                                                                                                                                                                                                                                                                                        rative
                                                                                                                                                                                                                                                                                                                                                              1.0
                                                                                                                                                                                                                                                                                                                                                                                                                                               .
.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     Š
                                                                                                                                                                                          13.
                                                                                                                                                                                                                                                                                                                                                                                                        ent.
                                                                     :::
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  Ξ
```

300 289

```
cardiovascular disease; immune system disorder; organ transplantation, tissue growth disorder; tissue regeneration disorder; diabetes mellitus; hypothyroidism; cholesterol ester storage disease; infection; vulnerary; vasotropic, entipsoriatic; antidiabetic; cytostatic; nootropic; no neuroprotective; antidatherosclerotic; anticogulant; thrombolyic; cardiant; hypotensive; antidhyroid; antidiflammatory; immunomodulator; dermatological; analgesic; virucide; antibacterial; fungloide.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  Human, ORF, cpen reading frame, ORFX, drug screening; diagnosis; disease monitoring; cytokine; cell proliferation; cell differentiation; immune modulation; haematopoiesis; requiation; tissue growth; anglogenesis; activin; inhibit; chemotactic; chemotactic; haemastatic; thrombolytic; tumour inhibition; bodily characteristic; fertility;
                                                                                                                                                                                                                                                                                                                                                                                                The present sequence is a polypeptide encoded by one of a large number of 5. ESTs derived from mRNAs encoding secreted proteins. The 5. ESTs different tissues. EST sequences usually correspond mainly to the 3. untranslated region (UTR) of the mRNA because they are often obtained from oligo-dT primed CDNA libraries. Such ESTs are not well suited for isolating cDNA sequences derived from the 5. ends of mRNAs and even those cases where longer cDNA sequences have been obtained, the full 5. UTR is rarely included. 5. ESTs are derived from mRNAs with intact 5. ends and can therefore be used to obtain full length cDNAs and genomic DNAs. 5. ESTs are also used in diagnostic, forensic, gene therapy and
                                                                                                                                                                                                                                                                            New nucleic acid that is a 5' expressed sequence tag (5' EST) for obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for diagnostic, forensic, gene therapy and chromosome mapping procedures
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        1 MARVEGAPVALGEWSLCWSLAIATPEPPTSABGNVAEGETKPDPOVTERCSDGWSFDAIT 60
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 1 MARVLGAPVALGLWSLCWSLAIATPLPPTSAHGNVAEGETKPDPPDVTERCSDGWSFDATT 60
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           chromosome mapping procedures. They are used to obtain upstream regulatory sequences and to design expression and secretion vectors.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    behaviour; cancer; proliferative disorder; neurological disorder
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      20.9%; Score 456; DB 21; Lungth 83; 98.8%; Pred. No. 3.5e-35; ive 0; Mismatches 1: Indels
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              Human glycoprotein-like ORF550 protein, SEG ID NO:1100.
                                                                                                                                                                        Giordano J:
                                                                                                                                                                                                                                                                                                                                                                Claim 13; SEQ ID 4385; 71pp + CD-POM: English
                                                                                                                                                                          Duclert A.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                ABP31577 standard; Protein; 87 AA.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   61 LDDNGTMLFFKGEXVWKSHKWDR 93
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 LDDNGTMLFFKGEFVWKSHKWDR 83
                                           21.FEB-2000; JUCOBE-0200610.
                                                                                      99US-0122487
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       09-JUL-2002 (first entry)
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    Conservative
                                                                                                                                                                          Dumas Milne Eiwards J,
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               Query Match
Best Local Similarity
                                                                                                                                                                                                                 2000-500381/45.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    83 AA;
                                                                                                                                                                                                                                        N-PSDB; AAC00310.
                                                                                                                               (GEST ) GENSET.
                                                                                      26-FEB-1999;
06-SEP-2000
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              ABP31577;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    Sequence
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 61
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            RESULT 3
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                Db
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 ŏ
```

;

Homo sapiens

Ė